

Zadatak R: Grupisanje zemalja po stepenu sreće 2017

Dati su podaci iz Izveštaja o sreći u svetu 2017 koji objavljuju Ujedinjene nacije svake godine na Dan sreće 20. marta. Izveštaj obuhvata 155 zemalja i uključuje faktore koji utiču na sreću: ekonomsku snagu, socijalnu podršku, očekivani životni vek, sloboda, odsustvo korupcije i velikodušnost. Dataset se nalazi u fajlu "world-happiness-report-2017.csv", a varijable dataset-a su:

- **Country** - naziv zemlje
- **Happiness Rank** - rang zemlje u 2017. godini
- **Happiness Score** - stepen sreće
- **Whisker high** - maksimalna vrednost stepena sreće date zemlje (ne računajući outliers)
- **Whisker low** - minimalna vrednost stepena sreće date zemlje (ne računajući outliers)
- **Economy GDP per Capita** - BDP po stanovniku
- **Family** - prosečan broj članova porodice
- **Health Life Expectancy** - očekivani životni vek
- **Freedom** - sloboda
- **Generosity** - velikodušnost
- **Trust Government Corruption** - poverenje u vladu i odsustvo korupcije
- **Dystopia Residual** - razlika između stepena sreće date zemlje i stepena sreće distopijske zemlje (imaginarne zemlje u kojoj žive najmanje srećni ljudi)

Zadatak je primenom **k-means** algoritma identifikovati grupe tj. klastere instanci.

Potrebno je:

- Potrebno je odabrati attribute koji će biti uključeni u model i navesti razloge za odabir, odnosno neodabir atributa.
- Pripremiti podatke za primenu algoritma. U analizi koristiti podskup podataka koji sadrži samo podatke za zemlje koje imaju vrednost varijable *Economy* veći od 0.
- Primenom *Elbow* metode utvrditi najbolju vrednost za broj klastera (k).
- Uraditi klasterizaciju za izabranu (tj. utvrđenu najbolju) vrednost za k.
- Interpretirati dobijene klastere (grupe) na osnovu: broja instanci po klasteru, centara klastera, disperzije od centra.